anand guru

SOC EXPERTS

Interview Questions on **SIEM**

Anand Guru

Security+ | CySA+ | CEH | ECIH

Founder

SOC Experts

https://socexperts.com

What is SIEM?

SIEM stands for Security Information and Event Management.

It is security management solution that helps in collecting, parsing and correlating events from various log sources

anand guru

Why do we need SIEM?

When we have security solutions like AV, Firewall, IPS why do we need SIEM?

- Various security solution that we use to protect our network and data work in isolation.
- However, in order to detect todays sophisticated attacks, it would be helpful if we could correlate information from various devices. This correlation is provided by SIEM.
- AV, F/W, IPS are all preventive technologies where as SIEM is mostly used for detection and analysis.
- We need SIEM for regulatory compliances too.

anand guru

What are the popular SIEM vendors?

- According to Gartner's Magic Quadrant the leaders in SIEM market are
  - ✓ Splunk Enterprise Security
  - ✓ IBM QRadar
  - ✓ Exabeam
  - ✓ Securonix
  - ✓ LogRhythm
  - ✓ RSA Security Analytics

- Other popular SIEM include
  - ✓ Microfocus Arcsight

Source: Gartner (February 2020)

What is Parsing?

- Parsing is process of converting unstructured data into structured format.
- It is during parsing where SIEM extracts all the useful metadata like Source IP, Destination IP, Username, File Name, File Hash etc. from the logs.

Raw Log **Jan 1 20:28:02 knight sshd[20336]: Failed password for root from 218.49.183.17 port 49869 ssh2**

Parsing

Parsed Event

Metadata

Event = Failed Password

Username = root

Source IP = 218.49.183.17

Source Port = 49869

Protocol = ssh

What is Normalization?

- Normalization is a process of bringing all events in to one common structure to deliver a homogeneous view
- It could be time normalization (bring events from various devices from different geo-location to a common time-zone)
- Identifying common event attributes (Like categorization – stamping all login, logout events as Authentication)

Please Note: For some SIEM tools **Normalization = Parsing**

What is Aggregation?

What is the use of aggregation?

- It is the process of merging similar logs that occur over a period of time.
- For example if there are 20 failed login events by a user on a server, the server would generate 20 logs. But, instead of storing all the 20 logs, SIEM will only store 1 record.
- Aggregation greatly helps in conserving the disk space and also increases the performance.

Raw Logs

Jan 1 20:28:25 Failed password for root from 218.49.183.17 port 49869 ssh2
Jan 1 20:28:37 Failed password for root from 218.49.183.17 port 49869 ssh2
Jan 1 20:28:41 Failed password for root from 218.49.183.17 port 49869 ssh2
Jan 1 20:28:49 Failed password for root from 218.49.183.17 port 49869 ssh2
Jan 1 20:28:57 Failed password for root from 218.49.183.17 port 49869 ssh2
Jan 1 20:29:02 Failed password for root from 218.49.183.17 port 49869 ssh2
Jan 1 20:29:13 Failed password for root from 218.49.183.17 port 49869 ssh2
Jan 1 20:29:20 Failed password for root from 218.49.183.17 port 49869 ssh2
Jan 1 20:29:35 Failed password for root from 218.49.183.17 port 49869 ssh2

Aggregation

Stored Record

Jan 1 20:28:02 Failed password for root from 218.49.183.17 port 49869 ssh2
Event Count = 9

What is Correlation?

- Correlation can be defined as set of conditions that indicates a suspicious activities.

  Examples:

  - User activity during non-business hours
  - VPN logins from multiple locations
  - Multiple malwares on the same host

anand guru

What are the different collections methods available in SIEM solution?

How do you on-board Windows Event Logs?

What are the log sources you have on-boarded?

Collection Method	Used for	Example Log Sources
WMI (Windows Management Instrumentation)	Windows Event Logs	Application, System and Security Logs
Syslog	Network Devices and Security Solutions	Routers, Switches, Firewall, Proxy, WAF etc.
Flat File CIFS (Windows) NFS (Linux)	Applications/Servers that write logs on to a flat file	 DNS, DHCP Apache Logs, Linux
Agent WinCollect – IBM QRadar SIEM Collector Agent – McAfee ESM Universal Forwards - Splunk	Various logs source including Windows Event Logs, Flat file logs etc.	Application, System and Security Logs, DNS, DHCP etc.
API	Cloud logs, Threat Feeds	AWS Cloudtrail logs
DB Collectors (JDBC/ODBC)	Databased Logs	MySQL Audit Logs, McAfee ePO, Oracle DB Audit Logs etc.

What are pull and push collection methods?

- A pull method of collecting logs means, the SIEM is actively involved in collecting the logs, usually by logging in to the log source. This is done at regular frequency.

  Example: WMI, Flat File, DB Collectors, API

- In push method, the logs are pushed by the log sources and SIEM just listens on the assigned port and IP.

  Example: Syslog, Agent

Pull and Push are sometimes referred to as **Active** and **Passive** Collection respectively

Syslog Collection Method.

1. Default port for Syslog?
    - Syslog – 514
    - TLS Syslog – 6514
2. Is Syslog TCP or UDP?
    - It has both TCP and UDP version

anand guru

While purchasing a SIEM, how do you size it?

How will you decide the capacity of SIEM you need?

- SIEM sizing is done in 2 ways – Events Per Second (EPS) and Data Indexed per day
- There are few methods to do this:
  1. We can set up a POC (proof of Concept) and onboard few log sources of each type and measure the EPS. Later we can multiply by appropriate number of devices.

     Example: If we had 200 Windows log source and one windows log source during POC generated 10 EPS, we will need 200 x 10 = 2000 EPS
  2. Easier method would be to use the vendor provided spreadsheet calculator and feed in all the data like, no. of windows log source, number of DNS, no. of R&S, no. of firewalls etc. This will give an approximate EPS. To be on safer side we should keep a extra buffer of 20-30%
- In order to calculate the amount of data indexed most SIEM vendors typically take 1 event is roughly equal to 400 bytes for many log source and 1kb for Windows Event Logs

  If we have 300 log sources that generate 10 EPS each, then EPS = 3000 and the amount of data indexed will roughly be 3000 x 400 (bytes) = 1,200,000 bytes/second

  1,200,000 x 60 x 60 x 24 = 103,680,000,000 ~= 103 GB per day

How long should we retain logs in SIEM?

- The retention is usually defined by the regulatory compliance. Like, PCI defines that (raw) logs must be stored for one year with the last three months available in an easily accessible storage.
- Parsed events should be retained for 90 days.

  Parsed events are typically used for analysis and reporting. It is very rare that we do analysis of events older than 90 days.
- Raw logs can be stored for an year.

anand guru

How Frequently should you take SIEM backup?

- It is recommended to schedule the SIEM backup for every 7 days.
- But we should take manual backup every time we are making major changes or upgrades

anand guru

How do you troubleshoot if any log source is not sending the logs?

- The non reporting logs sources are called as Silent Log Sources
- To troubleshoot a silent log source, we need to identify where the issue exists in the log flow. A typical log flow includes

- It could be at Log Source side - misconfiguration (wrong IP or port, necessary services not running, incorrect logging level etc.)
- The issue could be at network level - We can identify this by running packet capture on the collector (TCPDUMP for Linux based collectors and Wireshark for windows based) **`tcpdump –vvnni eth0 host <ip_logsource>`**
- The issue could be with collecting and parsing. We need to check appropriate services are running and check the logs files for any errors in SIEM
- Finally it could be a SIEM DB issue. which means the logs are being collected, parsed and stored, but UI is unable to query the DB. Usually a reboot would fix such issue.
- If none of this helps, I would bring in the vendor TAC

What is List/Watchlist/Reference Sets?

List is a collection of similar type of elements (like a list of IP addresses, list of all admin users). The lists can be used in filters and correlation rules.

Example:

1. List of all admin users
2. List of all Public IPs of a company
3. List of all Service Accounts

Lists are heavily used to integrate Threat Intelligence with SIEM

anand guru